Revised: EMI-2022-0786

Environmental Microbiology

(submitted 07/18/2022)

SUPPORTING INFORMATION

Ecophysiological and genomic analyses of a representative isolate of highly abundant *Bacillus cereus* strains in contaminated subsurface sediments

Jennifer L. Goff[1], Elizabeth G. Szink[1], Michael P. Thorgersen[1], Andrew D. Putt[3], Yupeng Fan[4], Lauren M. Lui[2], Torben N. Nielsen[2], Kristopher A. Hunt[5], Jonathan P. Michael[4], Yajiao Wang[4], Daliang Ning[4], Ying Fu[4], Joy D. Van Nostrand[4], Farris L. Poole II[1], John-Marc Chandonia[2], Terry C. Hazen[3,6,7], David A. Stahl[5], Jizhong Zhou[4,8,9,10], Adam P. Arkin[2,11], and Michael W.W. Adams[1,†]

[1]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA
[4]Insitute for Environmental Genomics, University of Oklahoma, Norman, OK, USA
[5]Civil and Environmental Engineering, University of Washington, Seattle, WA, USA
[6]Genome Sciences Division, Oak Ridge National Lab, Oak Ridge, TN, USA
[7]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA
[8]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA
[9]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA
[10]Earth and Environmental Sciences, Lawrence Berkley National Laboratory, Berkeley, CA, USA
[11]Department of Bioengineering, University of California at Berkeley, Berkeley, CA, USA
[†]Address correspondence to adamsm@uga.edu

Corresponding Author:
Michael W. W. Adams
(ORCID ID: 0000-0002-9796-5014)
Department of Biochemistry & Molecular Biology
University of Georgia
Athens, GA 30602, USA
Email: adamsm@uga.edu
Phone: 706 542 2060

TABLE OF CONTENTS

Figure S1. CPTF-Matching ASV Alignment……...3

Figure S2. Top ten ASVs in the subsurface sediment survey……...….4

Figure S3. Distribution of IS elements across the strain CPTF genome……..5

Figure S4. Growth of strain CPTF with nitrite……6

Figure S5. Evolution of the plasmid-borne ArsR-encoding regions ……..……7

Figure S6. Growth of strain CPTF with individual metals.....……8

Figure S7. Nitrate reduction during growth with COMM**.** …….……...9

Table S1. *B. cereus* metadata…….see attached Excel spreadsheets

Table S2. Plasmid content of strain CPTF……...……10

Table S3. Predicted IS elements in the strain CPTF genome……..11

Table S4. *B. cereus* IS elements…….see attached Excel spreadsheets

Table S5. Pangenome matrix……..see attached Excel spreadsheets

Table S6. Metal resistance genes……...see attached Excel spreadsheets

Table S7. Contaminated ORR metals mix (COMM) composition……12

Table S8. Genes used for *B. cereus* multi-locus sequence alignment ……...13

Table S9. Relative abundance of CPTF-matching ASV in ORR Area 3 groundwater ……14

Supplemental Discussion……....15

References……...…16

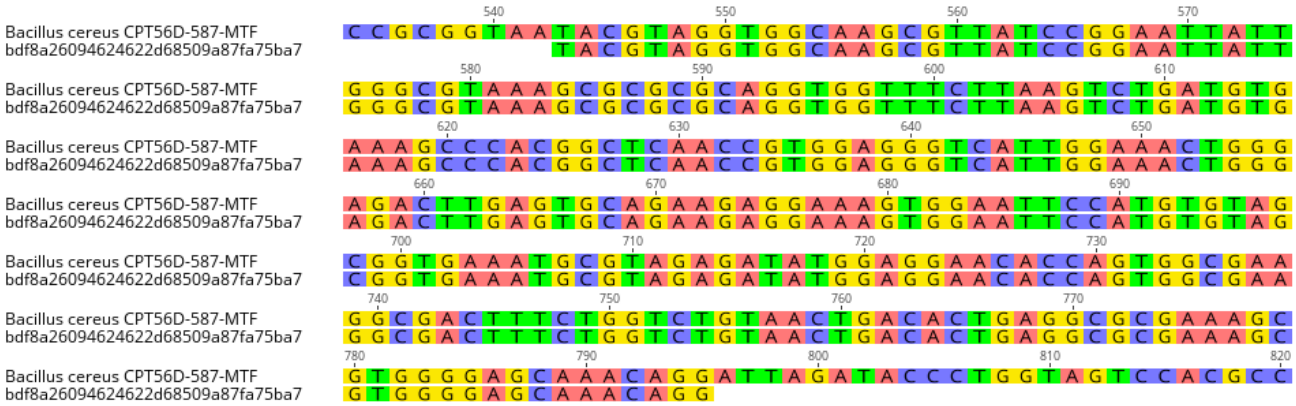

Figure S1. CPTF-Matching ASV Alignment. Alignment of the CPTF 16S sequence with the ORR 16S ASV bdf8a26094624622d68509a87fa75ba7 was performed with MUSCLE 3.8.425 implemented in Geneious Prime v2021.2.

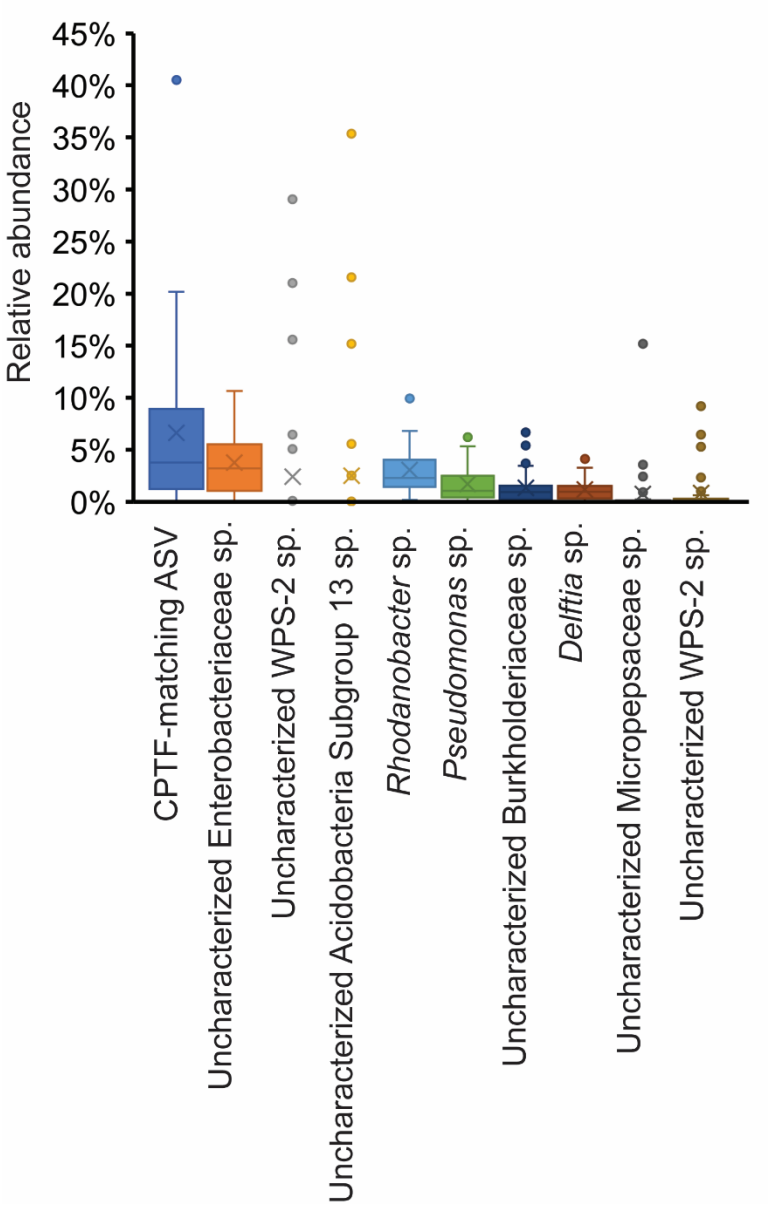

Figure S2. Top ten ASVs in the subsurface sediment survey. Overall relative abundance (displayed in order from left to right on the figure) was determined by summing all reads across all samples for each individual ASV. The summed read counts were then ranked, and the top ten results are shown here. The box-and-whisker plots display the relative abundance of the individual ASVs across all samples. The “X” indicates the mean relative abundance. Points represent outliers. The most detailed taxonomic assignment possible for the individual ASVs is displayed on the x-axis.

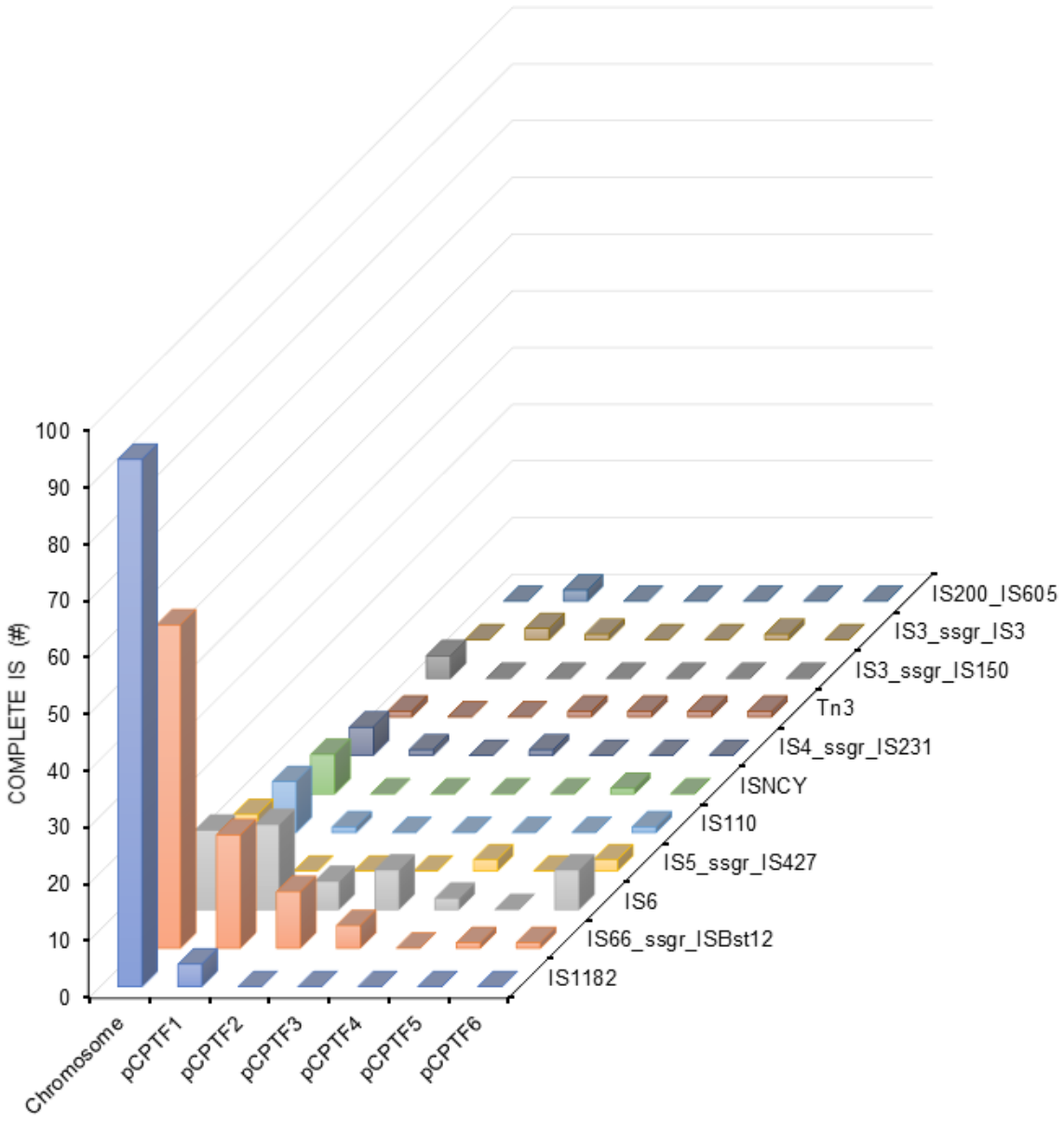

Figure S3. Distribution of IS elements across the strain CPTF genome. Genome features are shown along the x-axis. IS families are shown along the y-axis. “ssgr” indicates a subgroup of the IS family. The y-axis indicates the total number of complete IS elements within that family or family subgroup within a genomic region. Complete IS elements were determined using ISsaga.

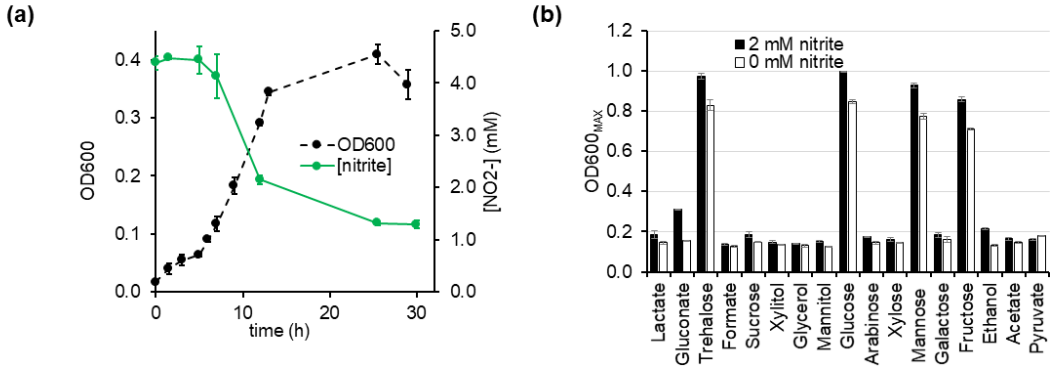

Figure S4. Growth of strain CPTF with nitrite (a) Nitrite reduction by strain CPTF. Strain CPTF is grown with nitrite (5 mM) and glucose (20 mM). Growth was determined by monitoring OD_{600} (black circles). Nitrite (green circles) concentrations were also measured. Individual circles represent the average of three biological replicates and error bars represent ±SD **(b)** To determine if nitrite reduction represents fermentative ammonification, growth of strain CPTF was monitored with different fermentable and non-fermentable carbon sources with (black bars) or without (white bars) added nitrite (2 mM). 20 mM of each respective carbon source was used. OD_{600} > 0.2 is scored positive for growth. Individual bars represent the average of three biological replicates and error bars represent ±SD. Differences in the extent of growth between panel **(a)** and **(b)** are due to the different light paths used for OD_{600} measurements between the two experiments.

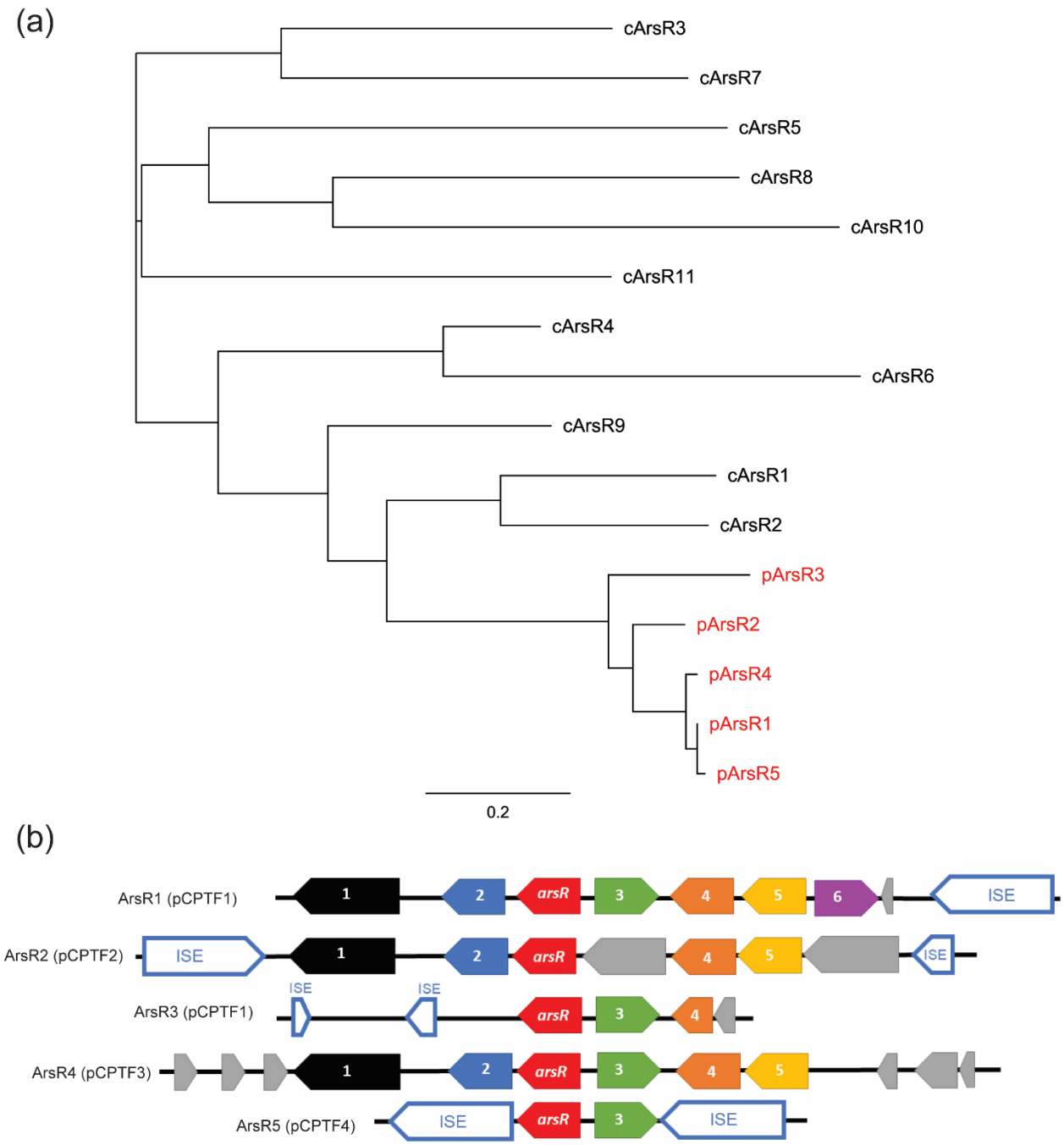

Figure S5. Evolution of the plasmid-borne ArsR-encoding regions. (a) Phylogenetic tree of 16 CPTF ArsR protein sequences. Alignment was performed using the Geneious Alignment algorithm with default parameters. The Neighbor-Joining tree was constructed using the Geneious Tree Builder. Plasmid ArsR sequences are highlighted in red. **(b)** Genetic regions of the five plasmid-borne *arsR* genes. ISE standards for insertion sequence element. CDSs encoding hypothetical proteins are colored in grey. 1 is a putative two-component sensor histidine kinase. 2 is the ribosomal protein L30e. 3 is RNA-binding protein Hfq. 4 is DNA-binding protein HBsu. 5 is a putative AbrB transcriptional regulator. 6 is an aspartyl-phosphate phosphatase Spo0E family protein.

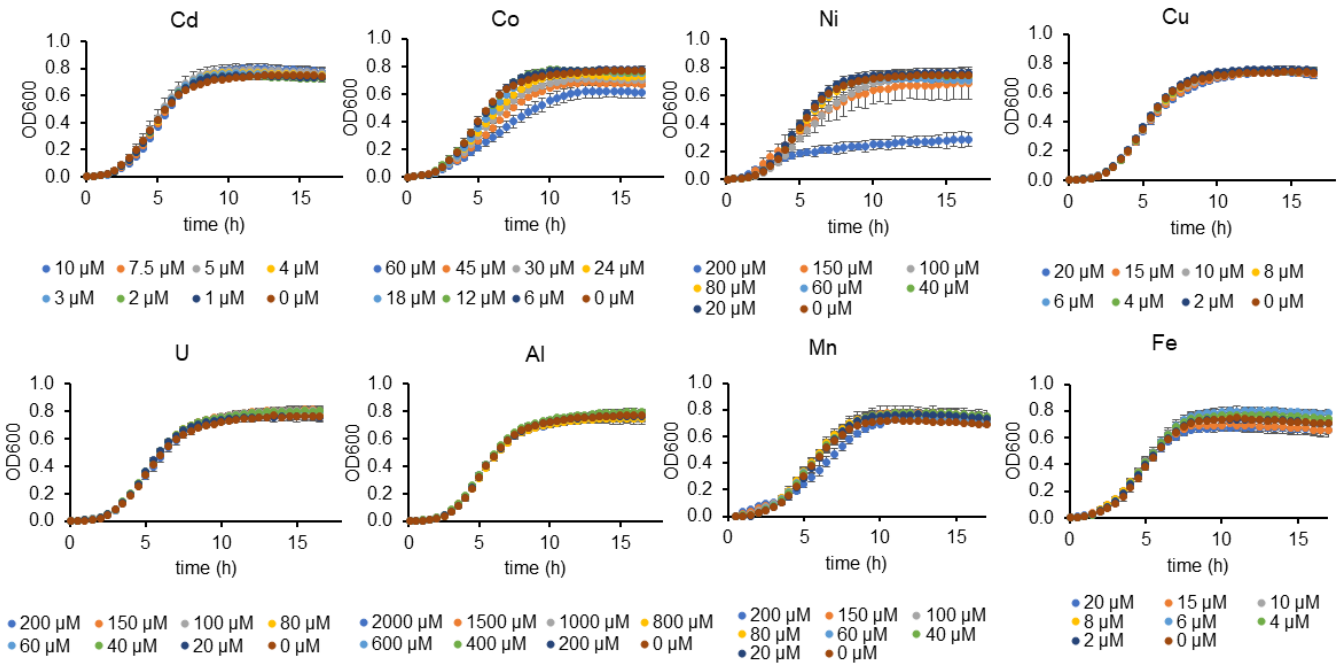

Figure S6. Growth of strain CPTF with individual metals. Metals were tested at site-relevant concentrations and represent 0.2-2X dilutions of the individual components of the COMM described in **Table S7.** All cultures were amended with 20 mM glucose and 10 mM nitrate and grown under anoxic conditions. Individual circles represent the average of three biological replicates and error bars represent ±SD.

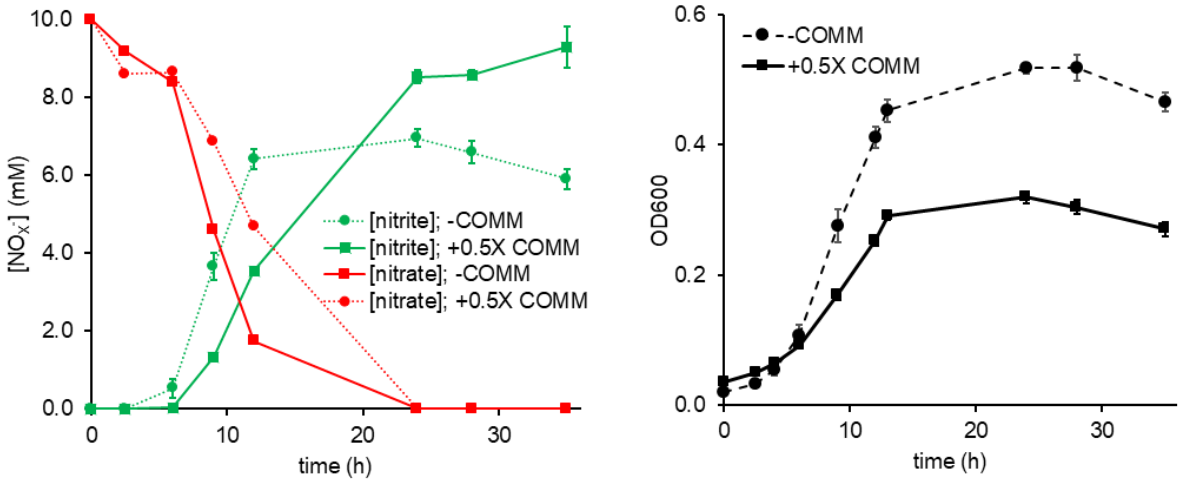

Figure S7. Nitrate reduction during growth with COMM. (a) Nitrate (red) and nitrite (green) concentrations during the growth of strain CPTF in the presence (solid line, square markers) and absence (dotted line, circle markers) of 0.5X COMM. 0.5X COMM concentrations are given in **Table S7**. Experiments were performed with glucose (20 mM) and nitrate (10 mM) under anoxic conditions. **(b)** Accompanying growth data to panel (a). Growth of strain CPTF in the presence (solid line, square markers) and absence (dotted line, circle markers) of 0.5X COMM For both plots, individual points represent the average of three biological replicates and error bars represent ±SD.

Table S2. Plasmid content of strain CPTF

Plasmid	Size (bp)	Copy number	GC content (%)	IS element content (%)[2]	Predicted plasmid type[1]
pCPTF1	448,451	2.46	34.20	19.07	Mobilizable
pCPTF2	99,643	2.23	34.95	29.00	Mobilizable
pCPTF3	83,506	2.96	34.25	33.78	Non-mobilizable
pCPTF4	83,382	2.10	34.95	10.68	Conjugative
pCPTF5	72,739	2.94	35.55	12.76	Non-mobilizable
pCPTF6	70,091	2.28	35.48	30.21	Non-mobilizable
pCPTF7	15,910	34.05	33.36	0.00	Non-mobilizable
pCPTF8	5,876	11.08	35.07	0.00	Non-mobilizable

[1]Plasmids were classified using the system described by Douarre et. al. (2020). [2]IS elements were predicted using ISsaga. The % IS content was calculated using both partial and complete IS elements.

Table S3. Predicted IS elements in the strain CPTF genome

ISFamily	IS Subgroup[1]	Complete[2]	Partial[2]
IS110	-	11	2
IS1182	-	97	101
IS200/IS605	-	2	20
IS200/IS605	IS1341	0	10
IS21	-	0	1
IS3	-	0	2
IS3	IS150	4	4
IS3	IS3	4	3
IS4	IS231	7	16
IS5	IS427	14	0
IS6	-	50	17
IS607	-	0	2
IS630	-	0	1
IS66	ISBst12	92	0
ISNCY	-	8	0
Tn3	-	5	1

[1]Subgroup given when identifiable by ISsaga. [2]Completeness determined by ISSaga

Table S7. Contaminated ORR metals mix (COMM) composition[1]

Metal	[ORR GW Median] (μM)[2]	[0.5X COMM] (μM)
Al	3385 (215-20700)	500
U	110 (40-576)	50
Mn	1202 (446-3150)	50
Ni	59 (19-157)	50
Co	8 (2-30)	15
Cu	1 (0.2-15)	5
Fe	1 (0.6-9.9)	5
Cd	1 (0.6-10)	2.5

[1]Described previously by Thorgersen et al. (2019). [2]Median values for six Area 3 wells. The range is given in parentheses.

Table S8. Genes used for *B. cereus* multi-locus sequence alignment

TIGRFAM ID	Gene name	TIGRFAM ID	Gene name
TIGR02673	FtsE	TIGR00048	rRNA_mod_RlmN
TIGR00019	prfA	TIGR00436	era
TIGR00635	ruvB	TIGR00165	S18
TIGR00484	EF-G	TIGR00431	TruB
TIGR00138	rsmG_gidB	TIGR02729	Obg_CgtA
TIGR00670	asp_carb_tr	TIGR00420	trmU
TIGR02493	PFLA	TIGR00020	prfB
TIGR00244	TIGR00244	TIGR00580	mfd
TIGR00150	T6A_YjeE	TIGR02386	rpoC_TIGR
TIGR00595	priA	TIGR00063	folE
TIGR01044	rplV_bact	TIGR00755	ksgA
TIGR03723	T6A_TsaD_YgjD	TIGR03632	uS11_bact
TIGR00767	rho	TIGR00223	panD
TIGR01389	recQ	TIGR03346	chaperone_ClpB
TIGR01083	nth	TIGR03263	guanyl_kin
TIGR00647	DNA_bind_WhiA	TIGR00430	Q_tRNA_tgt
TIGR00190	thiC	TIGR01744	XPRTase
TIGR00981	rpsL_bact	TIGR00612	ispG_gcpE
TIGR00447	pth	TIGR00184	purA
TIGR01169	rplA_bact	TIGR01455	glmM
TIGR00331	hrcA	TIGR01021	rpsE_bact
TIGR02350	prok_dnaK	TIGR03594	GTPase_EngA
TIGR00096	TIGR00096	TIGR01394	TypA_BipA
TIGR01163	rpe	TIGR00088	trmD
TIGR00510	lipA	TIGR00362	DnaA
TIGR00487	IF-2	TIGR00639	PurN
TIGR00485	EF-Tu	TIGR01050	rpsS_bact
TIGR00628	ung	TIGR01171	rplB_bact
TIGR01066	rplM_bact	TIGR00959	ffh
TIGR01855	IMP_synth_hisH	TIGR00468	pheS
TIGR01059	gyrB	TIGR01134	purF
TIGR01063	gyrA	TIGR01135	glmS
TIGR01060	eno	TIGR01073	pcrA
TIGR00563	rsmB	TIGR00482	TIGR00482
TIGR00042	TIGR00042	TIGR00036	dapB
TIGR02012	tigrfam_recA	TIGR00690	rpoZ
TIGR00382	clpX	TIGR01292	TRX_reduct
TIGR00521	coaBC_dfp	TIGR01255	pyr_form_ly_1
TIGR00065	ftsZ	TIGR01051	topA_bact
TIGR00064	ftsY	TIGR01011	rpsB_bact

Table S9. Relative abundance of CPTF-matching ASV in ORR Area 3 groundwater[1]

Area 3 Well ID	Relative Abundance (%)
FW106	0.004
FW104	0.007
FW109	1.345, 1.829, 1.485[2]
FW126	n.d.[3]
GW101	n.d.[3]

[1]Data is taken from Smith et al. (2015) [2]Different numbers represent samples taken from the same well at different time points. [3]not detected

SUPPLEMENTAL DISCUSSION

Additional genetic signatures of adaptation to sediment environment. Interestingly, the CPTF-matching ASV was either not detected or present only at low levels within the Area 3 groundwater samples pumped from on-site wells reported in a study by Smith *et al.* (2015). The median relative abundance of the CPTF-matching ASV in that groundwater study was 0.007% (range: n.d. to 1.829%) **(Table S9).** This is comparatively lower than its median relative abundance reported in the Area 3 sediment samples in this current study: 3.77% (range: 0.02 – 40.52%). Two non-mutually exclusive explanations for this observation are: (1) this ASV is present primarily in the sporulated lifestage in the groundwater but is not detected due to poor recovery of endospores during standard eDNA extraction (Filippidou *et al.*, 2015), or (2) this ASV represents a sediment-adapted microorganism that is ubiquitous across the subsurface sediment but not in the associated groundwater pumped from nearby wells. There is some genomic suggestion of the latter. As mentioned above, the genome encodes several adhesion factors which may facilitate microbial attachment to soil particle surfaces. These include surface adhesins which were found to be less common features of the *B. cereus* species (the five gene clusters containing the CPTF adhesins were also found in 19, 31, 34, 35, and 41 out of 42 genomes from the *B. cereus* pangenome analysis performed in this work) and flagellar structural and assembly proteins which are part of the *B. cereus* core/softcore genome. We found that the strain's plasmids encoded auxiliary genes that may be important for soil invertebrate host-microbe interaction (Moar *et al.*, 2017) and resource competition (Cytryn, 2013) within the ORR subsurface sediments including insecticidal endotoxins (Vachon *et al.*, 2012) and various antimicrobial peptides (Heilbronner *et al.*, 2021). For example, the pCPTF8 plasmid encodes a subtilisin-like precursor peptide (present in 5/42 genomes). Subtisilins can also function in maturation of lantibiotics and extracellular adhesins (Tripathi and Sowdhamini, 2008).

REFERENCES

Cytryn, E. (2013) The soil resistome: The anthropogenic, the native, and the unknown. *Soil Biol Biochem* **63:** 18-23.

Filippidou, S., Junier, T., Wunderlin, T., Lo, C.-C., Li, P.-E., Chain, P.S.,and Junier, P. (2015) Under-detection of endospore-forming Firmicutes in metagenomic data. *Computational and Structural Biotechnology Journal* **13:** 299-306.

Heilbronner, S., Krismer, B., Brötz-Oesterhelt, H.,and Peschel, A. (2021) The microbiome-shaping roles of bacteriocins. *Nat Rev Microbiol* **19:** 726-739.

Moar, W.J., Evans, A.J., Kessenich, C.R., Baum, J.A., Bowen, D.J., Edrington, T.C., Haas, J.A., Kouadio, J.-L.K., Roberts, J.K., Silvanovich, A., Yin, Y., Weiner, B.E., Glenn, K.C.,and Odegaard, M.L. (2017) The sequence, structural, and functional diversity within a protein family and implications for specificity and safety: The case for ETX_MTX2 insecticidal proteins. *J Invertebr Pathol* **142:** 50-59.

Smith, M.B., Rocha, A.M., Smillie, C.S., Olesen, S.W., Paradis, C., Wu, L., Campbell, J.H., Fortney, J.L., Mehlhorn, T.L., Lowe, K.A., Earles, J.E., Phillips, J., Techtmann, S.M., Joyner, D.C., Elias, D.A., Bailey, K.L., Hurt, R.A., Preheim, S.P., Sanders, M.C., Yang, J., Mueller, M.A., Brooks, S., Watson, D.B., Zhang, P., He, Z., Dubinsky, E.A., Adams, P.D., Arkin, A.P., Fields, M.W., Zhou, J., Alm, E.J., Hazen, T.C.,and Lindow, S.E. (2015) Natural Bacterial Communities Serve as Quantitative Geochemical Biosensors. *mBio* **6:** e00326-15.

Thorgersen, M.P., Ge, X., Poole, F.L., 2nd, Price, M.N., Arkin, A.P.,and Adams, M.W.W. (2019) Nitrate-Utilizing Microorganisms Resistant to Multiple Metals from the Heavily Contaminated Oak Ridge Reservation. *Appl Environ Microbiol* **85**.

Tripathi, L.P.,and Sowdhamini, R. (2008) Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics* **9:** 549.

Vachon, V., Laprade, R.,and Schwartz, J.L. (2012) Current models of the mode of action of Bacillus thuringiensis insecticidal crystal proteins: a critical review. *J Invertebr Pathol* **111:** 1-12.